Hybrid network traffic engineering system (HNTES)

Zhenzhen Yan, Chris Tracy, Malathi Veeraraghavan
University of Virginia and ESnet
April 23, 2012
zy4d@virginia.edu, ctracy@es.net, mvee@virginia.edu

Project web site: http://www.ece.virginia.edu/mv/research/DOE09/index.html

Thanks to the US DOE ASCR program office and NSF for
UVA grants DE-SC002350, DE-SC0007341, OCI-1127340
and
ESnet grant DE-AC02-05CH11231

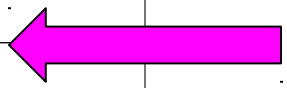
Problem statement

- A hybrid network supports both IP-routed and circuit services on:
  - Separate networks as in ESnet4, or
  - An integrated network as in ESnet5
- A hybrid network traffic engineering system (HNTES) is designed to move science data flows off the IP-routed network to circuits
- Problem statement: Design HNTES

Two reasons for using circuits

1. Offer scientists rate-guaranteed connectivity
2. Isolate science flows from general-purpose flows

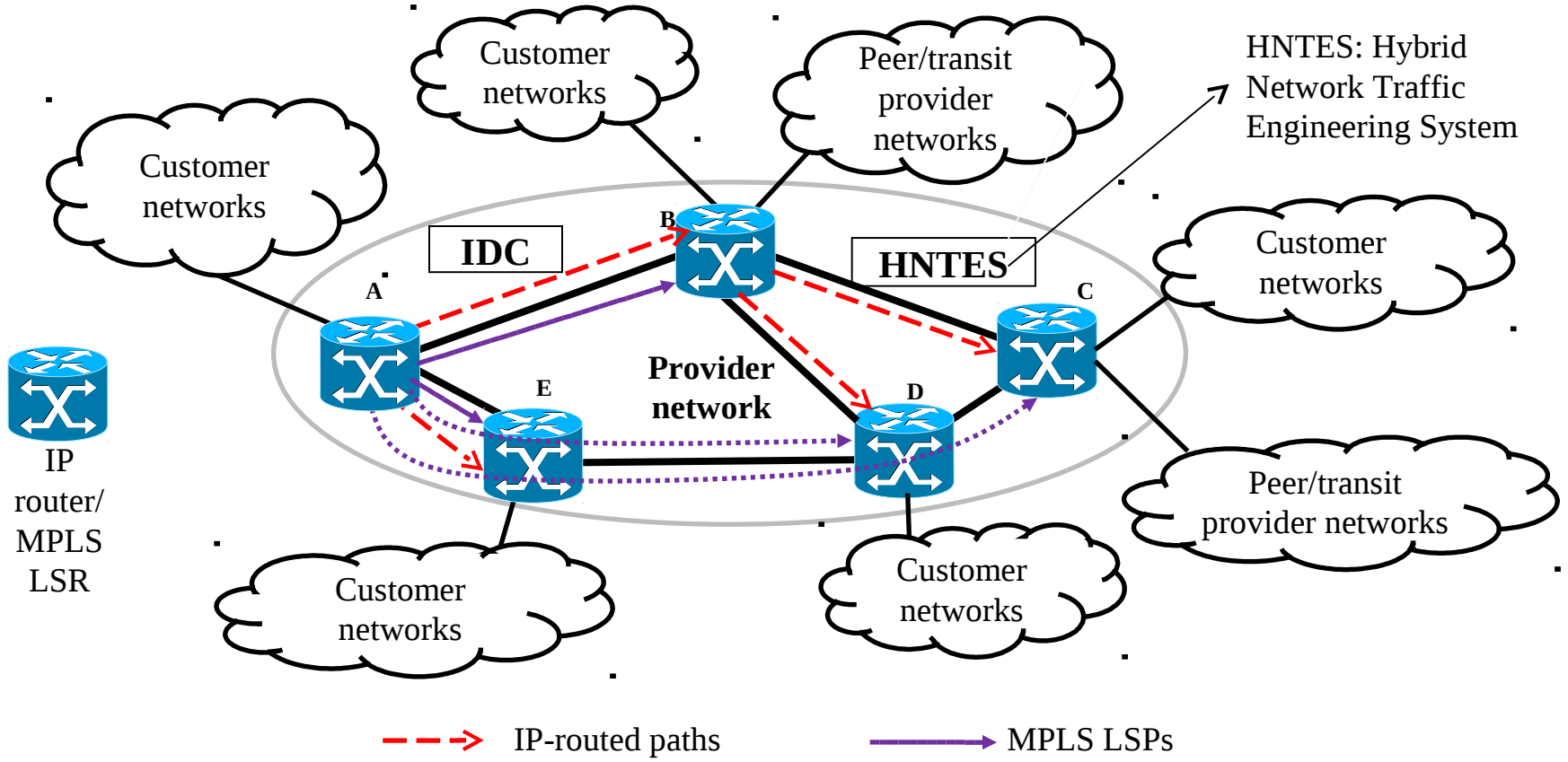
Reason / Circuit scope	Rate-guaranteed service	Science flow isolation
End-to-end (inter-domain)	✔	✔
Per provider (intra-domain)	✘	✔



Should we mine trouble ticket logs to quantify the negative impact of science flows on “beta” flows?

The “Why” question

Rest of the slides: Focus on the "How" question Usage within domains for science flow isolation



- Policy based routes added in ingress routers to move science flows to MPLS LSPs

HNTES Design questions

- What type of flows should be redirected off the IP-routed network?
- What are key components of a hybrid network traffic engineering system?
- Prove/disprove underlying hypothesis of design through ESnet NetFlow data analysis

First considered these options

- Dimensions
  - size (bytes): elephant and mice
  - rate: cheetah and snail
  - duration: tortoise and dragonfly
  - burstiness: porcupine and stingray

Kun-chan Lan and John Heidemann, A measurement study of correlations of Internet flow characteristics. *ACM Comput. Netw.* 50, 1 (January 2006), 46-62.

working answer

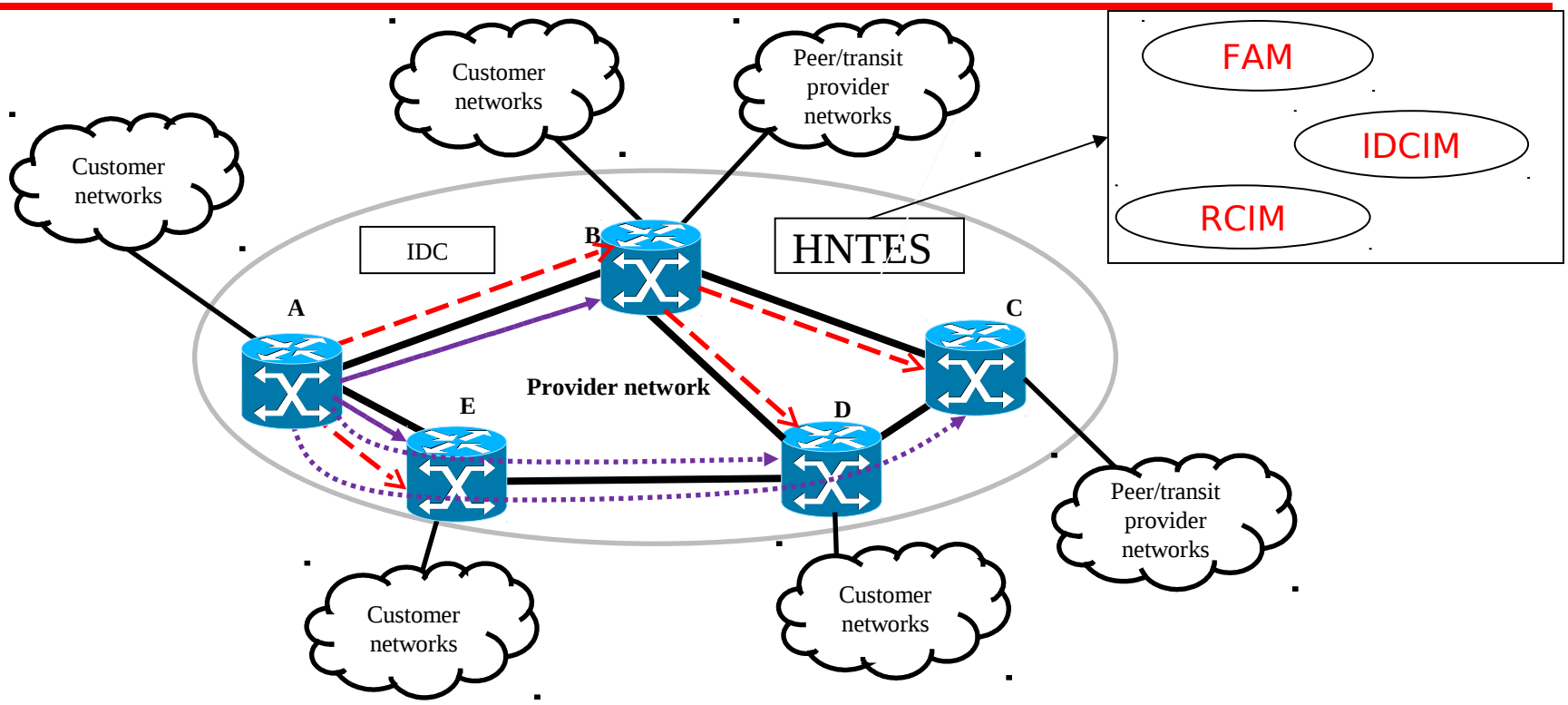
- alpha flows should be redirected
- what are alpha flows?
  - flows with high sending rates in any part of the lifetime
    - number of bytes in any T-sec interval $\geq$ H bytes
    - if H = 1 GB and T = 60 sec
      - throughput exceeds 133 Mbps
- alpha flows are
  - responsible for burstiness
  - caused by transfers of large files over high bottleneck-link rate paths
- who generates this type of flows?
  - scientists who move large sized datasets invest in high-end computers, high-speed disks, parallel file systems, and high access link speeds

S. Sarvotham, R. Riedi, and R. Baraniuk, “Connection-level analysis and modeling of nework traffic,” in ACM SIGCOMM Internet Measurement Workshop 2001, November 2001, pp. 99–104.

Design questions

- What type of flows should be redirected off the IP-routed network?
- ➢ What are key components of a hybrid network traffic engineering system?
- Prove/disprove underlying hypothesis of design through ESnet NetFlow data analysis

Components of HNTES

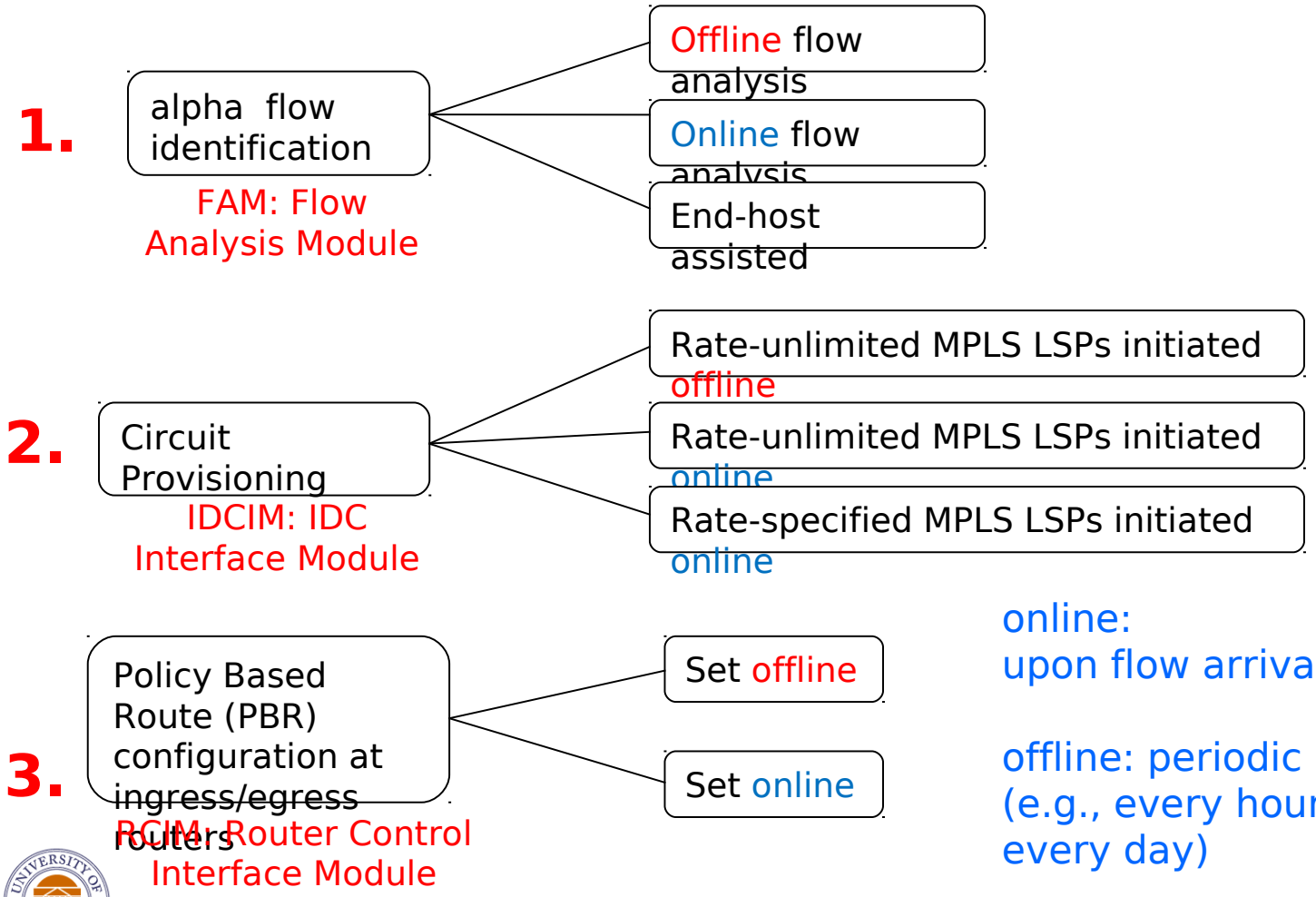


FAM: Flow Analysis Module

IDCIM: IDC Interface Module

RCIM: Router Control Interface Module

Three tasks executed by HNTES



online:
upon flow arrival

offline: periodic process (e.g., every hour or every day)

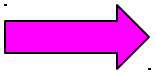
alpha flow identification

- Possible online methods
  - Method 1:
    - Today's routers support packet classification into flows and have the ability to measure rates (for rate policing)
    - But there is no mechanism for them to inform a management system when high-rate flows arrive
  - Method 2:
    - NetFlow: routers group packets into flows and send reports to a collector (files created at collector every 5 mins)
    - Raw netflow packets from the router can be collected by a host (or via a flow-fanout from current collector)
      - New flow information can be obtained every 60 sec (active timeout interval)
      - Identify high rate flows

online alpha flow identification methods contd.

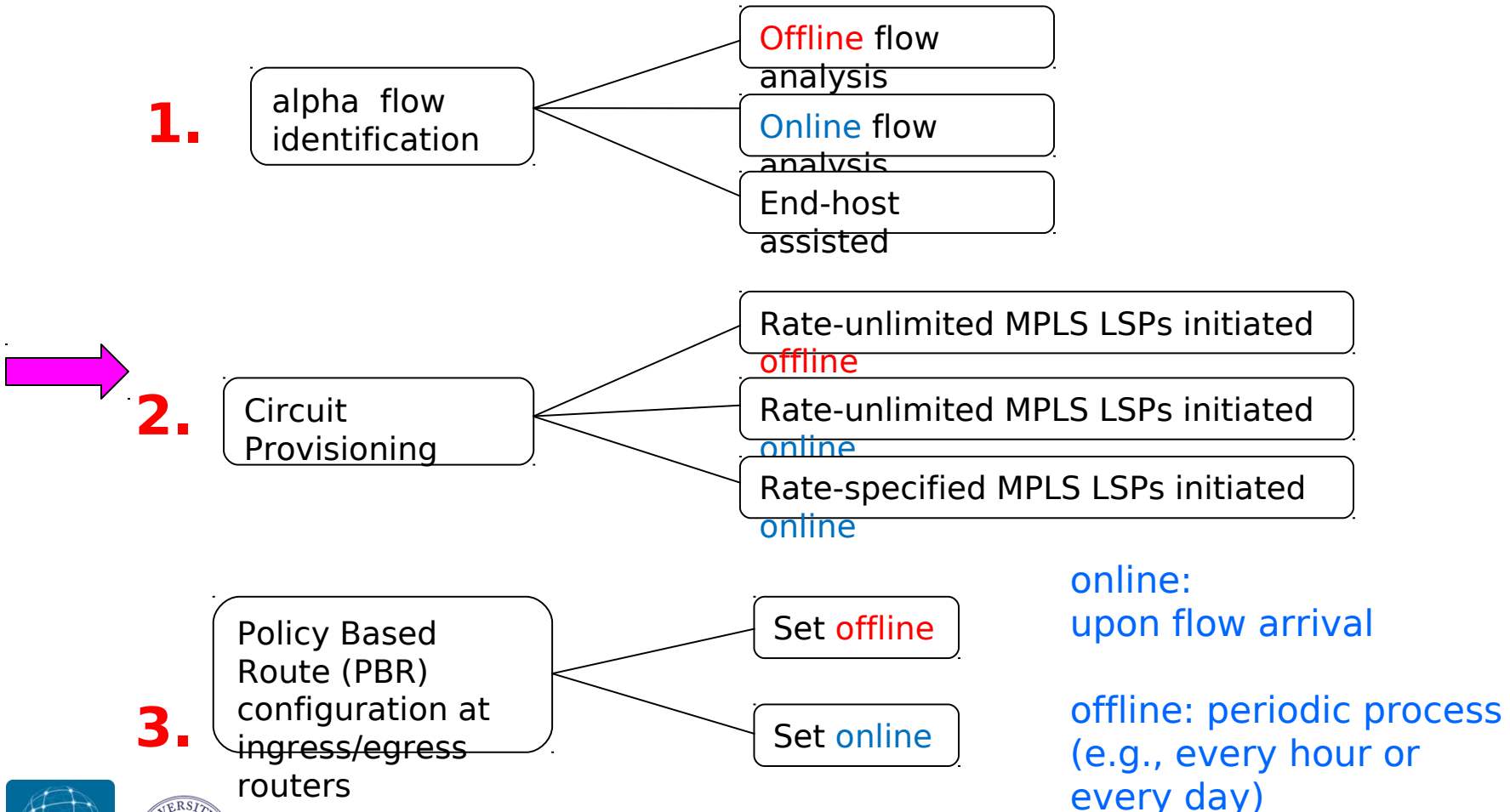
- Method 3:
  - Port mirror packets to external server and run algorithms to detect high-rate flows.
  - Cons: does not scale with link rate
    - May need many external servers
  - Deployment seems impractical: need a cluster of servers per ESnet router

Proposed solutions



- Solution 1
  - Strictly offline
  - Analyze NetFlow data on a daily basis and identify source/destination hosts (/32) or subnets (/24) that are capable of sourcing/sinking data at high rates → prefix flows
- Solution 2: Hybrid (NetFlow and Mirroring)
  - Combine offline scheme for /32 and /24 prefix flow ID, with
  - Online scheme
    - NetFlow with 10 sec reporting, OR
    - 0-length packet mirroring to external server for online detection of raw IP flows (5-tuple) whose IDs match offline configured prefix flow IDs

HNTES three tasks (revisit)



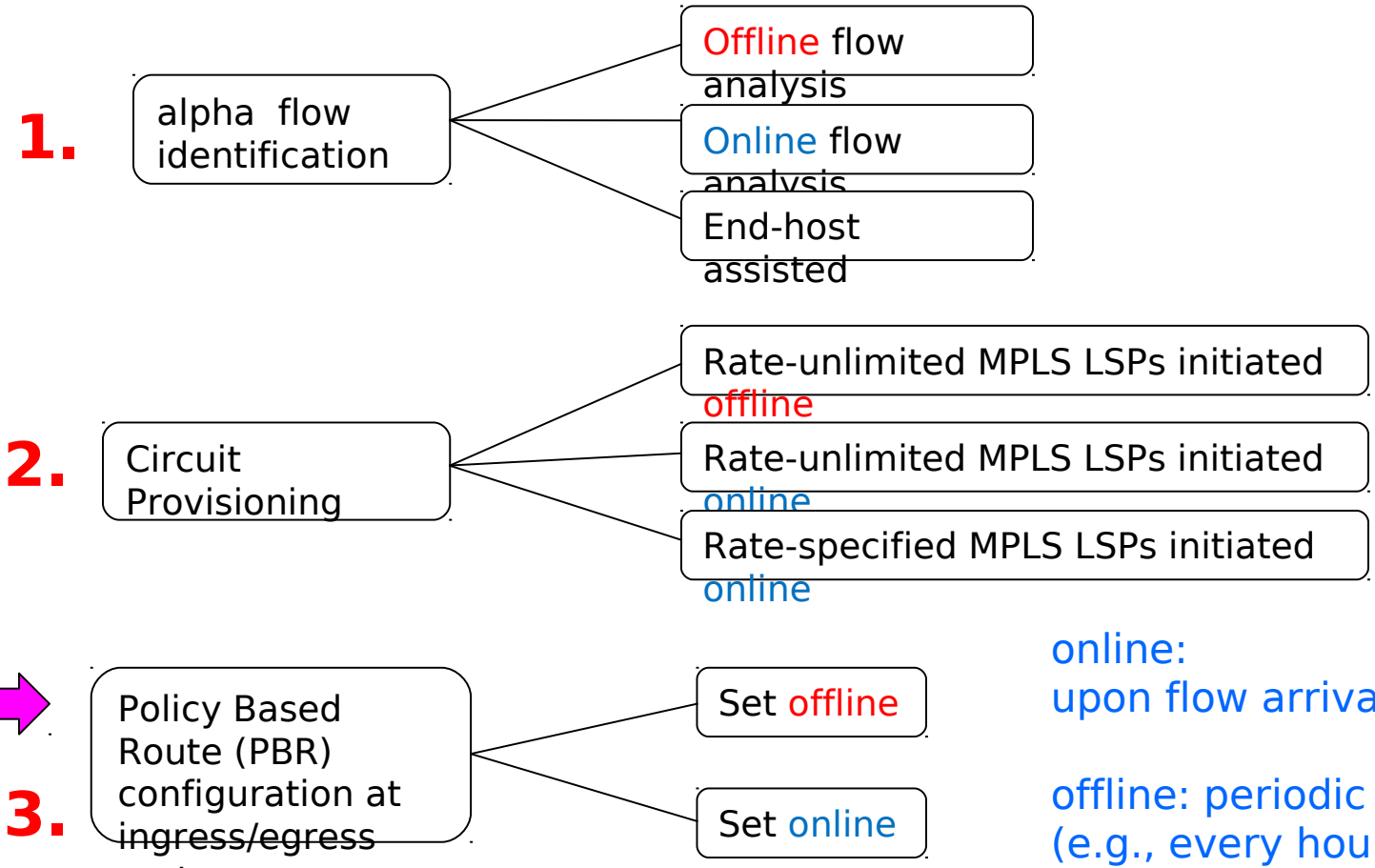
online:
upon flow arrival

offline: periodic process (e.g., every hour or every day)

Circuit Provisioning

- Circuits
  - rate-specified per-alpha flow specific circuits are desirable if goal is rate guarantee
  - but if circuits are only intra-domain with the purpose of isolating science flows, it is sufficient to configure routers to redirect multiple alpha flows to same rate-unlimited LSP
  - set up such LSPs a priori between all ingress-egress router pairs of provider's network that have seen alpha flows based on offline analysis

Three tasks executed by HNTES



online:
upon flow arrival

offline: periodic process (e.g., every hour or every day)

PBR configuration

- Online:
  - Commit operation in JunOS can take on the order of minutes based on the size of the configuration file
  - Sub-second configuration times for OpenFlow switches?
- Offline:
  - Cannot configure routes for 5 tuple raw IP flows as ports are ephemeral
  - Configuring PBRs for /32 or /24 prefix flows implies some beta flows will also be redirected to the science LSPs

HNTES design solutions

- All offline solution (discussed next)
- Hybrid online-offline solution
  - hybrid alpha flow identification
  - offline circuit provisioning
  - online PBR configuration for 5-tuple raw IP flows
- Pros/cons of hybrid scheme:
  - Pro: beta flows will not be redirected to VCs (avoid alpha flow effects)
  - Con: some alpha flows will end before redirection

Review of current (all offline) HNTES design

- Flow analysis module analyzes NetFlow reports on a daily basis (offline)
  - Prefix flow identifiers determined for subnets (/24) or hosts (/32) that can source-sink alpha flows
- Pairwise rate-unlimited LSPs provisioned between ingress-egress routers for which prefix flows were identified
- PBRs set at routers (both directions) for prefix flow redirection
  - Entries aged out of PBR table to keep it from growing too large

Design questions

- What type of flows should be redirected off the IP-routed network?
- What are key components of a hybrid network traffic engineering system?
- ➢ Prove/disprove underlying hypothesis of design through ESnet NetFlow data analysis

Hypothesis

- Key assumption in offline solution:
  - Computing systems that run the high-speed file transfer applications will likely have static public IP addresses, which means that prefix flow identifier based offline mechanisms will be effective in redirecting alpha flows.
  - Flows with previously unseen prefix flow identifiers will appear but such occurrences will be relatively rare

NetFlow data analysis

- NetFlow data over 7 months (May-Nov 2011) collected at ESnet site PE router
- Three steps
  - UVA wrote R analysis and anonymization programs
  - ESnet executed on NetFlow data
  - Joint analysis of results

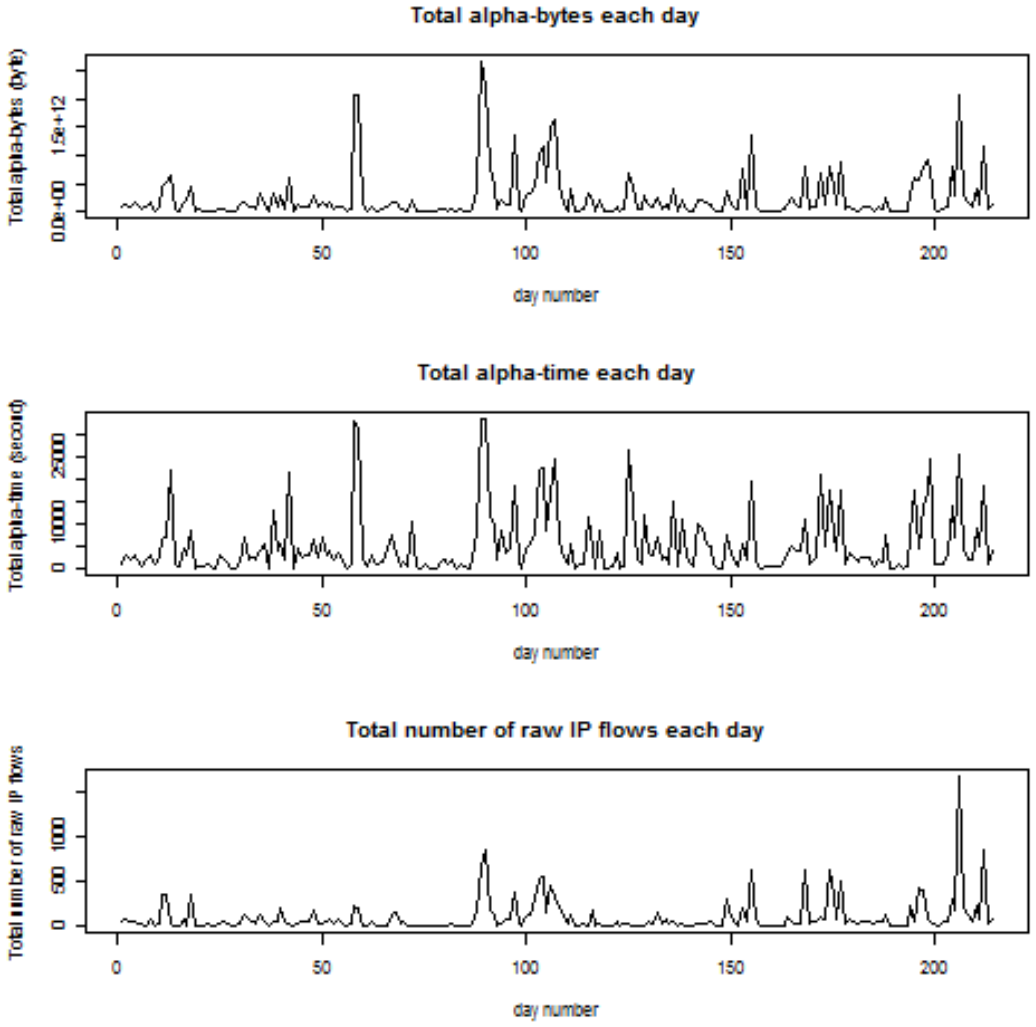
alpha flow identification algorithm

- alpha flows: high rate flows
  - NetFlow reports: subset where bytes sent in 1 minute > H bytes (1 GB)
  - Raw IP flows: 5 tuple based aggregation of NetFlow reports on a daily basis
  - Prefix flows: /32 and /24 src/dst IP aggregation on a daily basis
- Age out PBR entries
  - if for “A” aggregation intervals, no raw IP flows corresponding to a prefix flow appear

Analyses

- Analyses:
  - Characterize alpha flows
    - 22041 raw IP flows
    - 125 (/24) prefix flows
    - 1548 (/32) prefix flows
  - Study effectiveness of offline solution

Characteristics of alpha flows



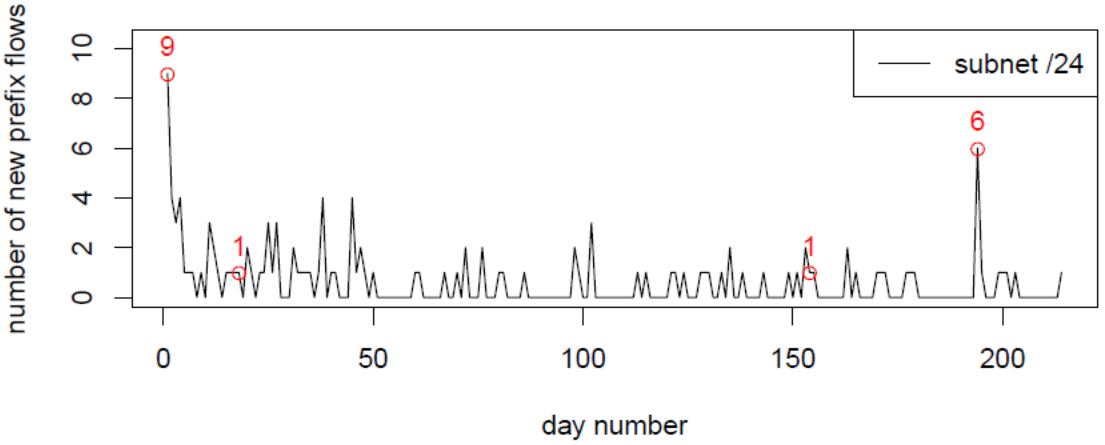
- Both alpha-bytes and alpha-time peaked on day 89
  - 2.65 TB
  - 9.3 hours
- Number of raw IP flows in a day:
  - One prefix flow had 1240 constituent alpha raw IP flows

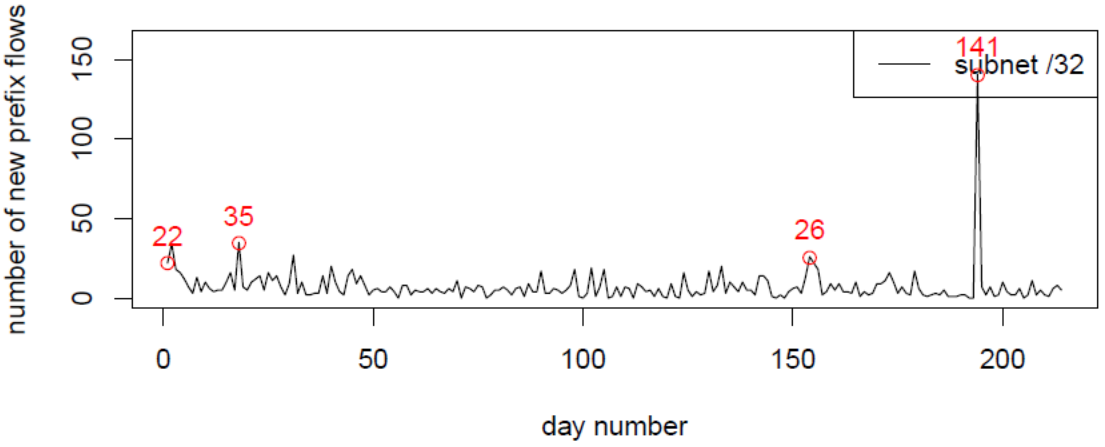
Two types of analyses

- Characterize alpha flows
- Study effectiveness of offline solution:
  - determine on a per-day basis, the percentage of bytes that came from flows that were not redirected because their prefix flow identifiers were not in the PBR table

Number of new prefix flows daily

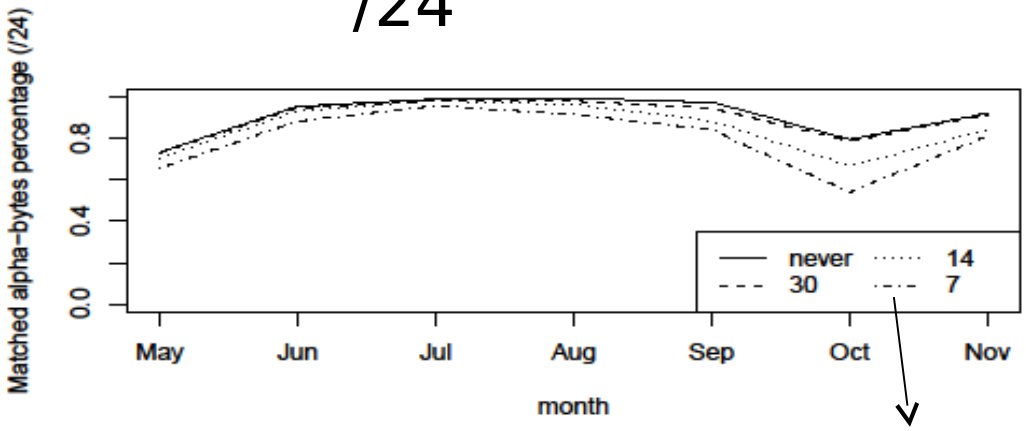
- For most days only 0 or 1 new prefix flow.
- When new collaborations start or new data transfer nodes are brought online, new prefix flows will occur





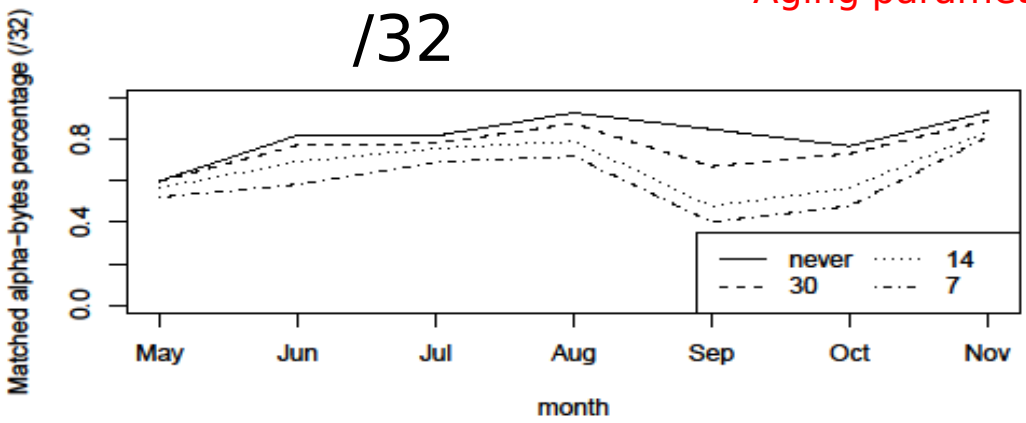
Percent of alpha bytes that would have been redirected

/24



Aging parameter

/32



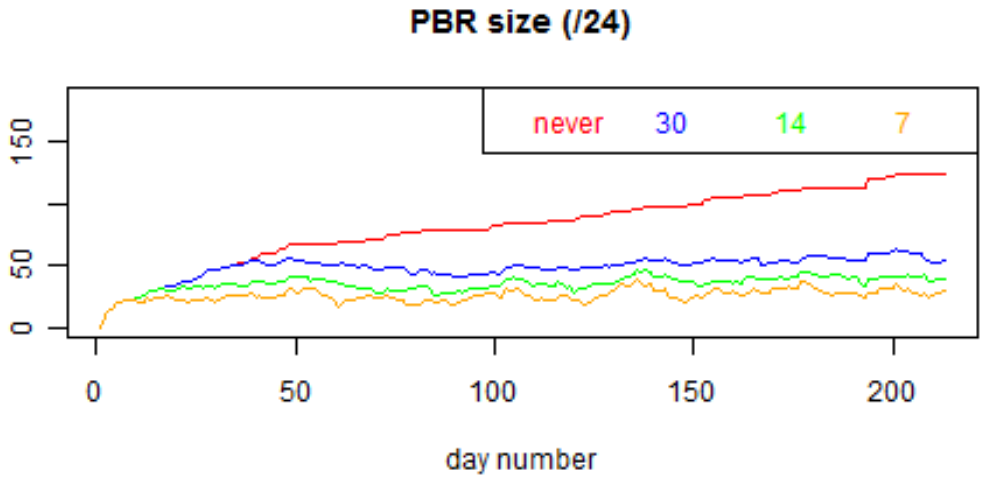
All 7 months:

Aging parameter	/24	/32
7	82%	67%
14	87%	73%
30	91%	82%
never	92%	86%

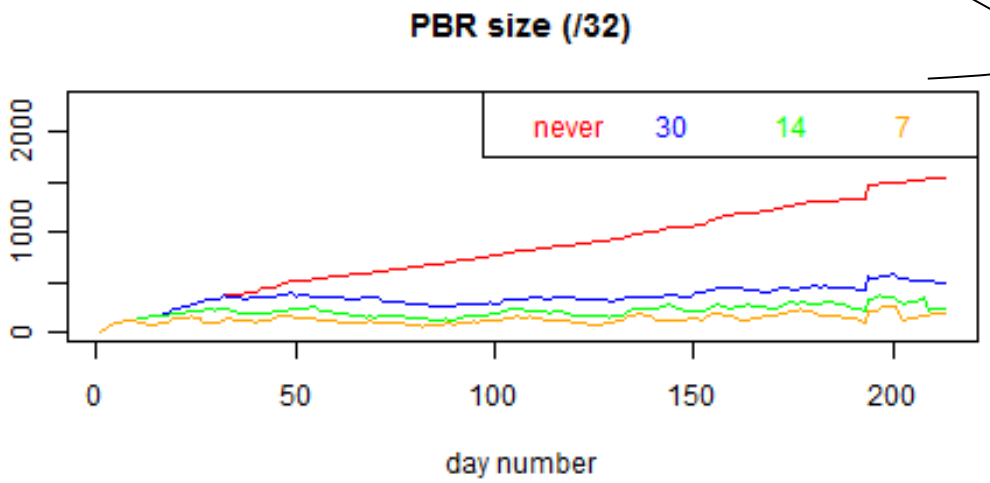
- When new collaborations start or new data transfer nodes are brought online, new prefix flows will occur, and so matched rates will drop

Effect of aging parameter on PBR table size

- For operational reasons, and forwarding latency, this table should be kept small
- With aging parameter =30, curve is almost flat



Aging parameter



Full mesh of LSPs required or just a few?

Number of super-prefix flows (ingress-egress router based aggregation of prefix flows) per month:

Month	May	Jun	July	Aug	Sep	Oct	Nov
total	13	15	16	16	18	18	18
repeated	0	13	15	16	16	18	18
new	13	2	1	0	2	0	0

Represents number of LSPs needed from ESnet site PE router to indicated numbers of egress routers

Conclusions

- From current analysis:
  - Hypothesis is true
  - Offline design appears feasible
    - IP addresses of sources that generate alpha flows relatively stable
    - Most alpha bytes would have been redirected in the analyzed data set
  - /24 seems better option than /32
  - 30 days aging parameter seems best: tradeoff of PBR size and effectiveness

Ongoing work

- NetFlow analyses
  - other routers' NetFlow data
  - quantify redirected beta flow bytes which will experience competition with alpha flows
  - utilization of MPLS LSPs
  - multiple simultaneous alpha flows on same LSPs
  - match with known data doors
- ANI testbed experiments
  - Out of order packets when PBR added
  - OpenFlow
  - Rate-unlimited LSPs
- Other HNTES designs
  - Hybrid design
  - End-application assisted design (Lambdastation, Terapaths)